

*Citation for published version:*

Dereich, S, Mailler, C & Morders, P 2017, 'Non-extensive condensation in reinforced branching processes', *Annals of Applied Probability*, vol. 27, no. 4, pp. 2539-2568. <https://doi.org/10.1214/16-AAP1268>

*DOI:*

[10.1214/16-AAP1268](https://doi.org/10.1214/16-AAP1268)

*Publication date:*

2017

*Document Version*

Peer reviewed version

[Link to publication](#)

## University of Bath

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Non-extensive condensation in reinforced branching processes

Steffen Dereich, Cécile Mailler and Peter Mörters

## Abstract

We study a class of branching processes in which a population consists of immortal individuals equipped with a fitness value. Individuals produce offspring with a rate given by their fitness, and offspring may either belong to the same family, sharing the fitness of their parent, or be founders of new families, with a fitness sampled from a fitness distribution  $\mu$ . Examples that can be embedded in this class are stochastic house-of-cards models, urn models with reinforcement, and the preferential attachment tree of Bianconi and Barabási. Our focus is on the case when the fitness distribution  $\mu$  has bounded support and regularly varying tail at the essential supremum. In this case there exists a condensation phase, in which asymptotically a proportion of mass in the empirical fitness distribution of the overall population condenses in the maximal fitness value. Our main results describe the asymptotic behaviour of the size and fitness of the largest family at a given time. In particular, we show that as time goes to infinity the size of the largest family is always negligible compared to the overall population size. This implies that condensation, when it arises, is non-extensive and emerges as a collective effort of several families none of which can create a condensate on its own. Our result disproves claims made in the physics literature in the context of preferential attachment trees.

**Keywords:** Network; preferential attachment; Bianconi-Barabási model; genetics; house-of-cards model; selection and mutation; urn model; reinforcement; non-Malthusian branching; Crump-Mode-Jagers process; condensation.

**MSC Classification:** Primary 60J80; Secondary 05C82; 60G70.

## Contents

<b>1</b>	<b>Background and motivation</b>	<b>2</b>
<b>2</b>	<b>Statement of the results</b>	<b>4</b>
<b>3</b>	<b>Proof of the condensation phenomenon</b>	<b>7</b>
3.1	The standard construction of the model . . . . .	8
3.2	General branching process theory . . . . .	9
3.3	A coupling technique . . . . .	9
<b>4</b>	<b>Estimates for the number of families in the population</b>	<b>10</b>
<b>5</b>	<b>Local convergence</b>	<b>11</b>
<b>6</b>	<b>Negligibility of families outside the main window</b>	<b>14</b>
6.1	Contribution of the unfit families . . . . .	15
6.2	Contribution of the families born late . . . . .	16
6.3	Families born early are not fit enough . . . . .	17
<b>7</b>	<b>Proof of non-extensiveness of condensation</b>	<b>18</b>
7.1	Proof of Theorem 2.2 . . . . .	18
7.2	Proof of Corollary 2.3 . . . . .	18
7.3	Proof of Theorem 2.4 . . . . .	19
<b>8</b>	<b>Open problems</b>	<b>20</b>

# 1 Background and motivation

The principal aim of this paper is to study the emergence of a condensate in stochastic models. For this purpose we consider a class of branching processes with reinforcement, which probably constitute the easiest class of models, in which this question can be studied in a meaningful way. Still we shall see that, due to the reinforcement, these models display rather complex behaviour and not all relevant questions on their behaviour will be answered.

Although our models can describe a variety of objects, see the examples below, we shall describe them as a structured population. Parameters of our model are a fitness distribution  $\mu$  on the positive reals, and positive numbers  $\beta, \gamma \leq 1$  with  $\beta + \gamma \geq 1$ . At any time  $t$  the population consists of a finite number  $N(t)$  of individuals. Each individual in the population has a fitness, and individuals are organised into families, such that all members of a family have the same fitness. The process is started with one family of one individual, whose fitness is drawn from the distribution  $\mu$ . Suppose, at time  $t \geq 0$ , the population consists of  $M(t)$  families, and there are  $Z_n(t)$  individuals of fitness  $F_n$  in the  $n$ th family, for  $1 \leq n \leq M(t)$ . Independently in every family birth events occur with a time-dependent rate  $F_n Z_n(t)$ . When a birth event occurs in the  $n$ th family, independently of everything else, one or both of the following happen,

- with probability  $\beta$  a new family is founded, initially consisting of one individual equipped with a fitness drawn, independently of everything else, from the distribution  $\mu$ ;
- with probability  $\gamma$  a new individual with fitness  $F_n$  is added to the  $n$ th family. The ability of the system to reproduce particles of the same type constitutes the reinforcement, see [21].

Note that both things happen simultaneously with probability  $\beta + \gamma - 1 \geq 0$ . If  $\mu$  has all exponential moments the total number  $N(t)$  of individuals in the population remains finite at all times, see for example Corollary 3.3 in [19]. Our main object of interest is the *empirical fitness distribution* at time  $t$ , which is defined as

$$\Xi_t = \frac{1}{N(t)} \sum_{n=1}^{M(t)} Z_n(t) \delta_{F_n}. \quad (1)$$

In this paper we focus on bounded fitness distributions  $\mu$ , and specifically the case in which a condensation phenomenon occurs, which we describe in some detail. Different phenomena occur in the case of unbounded fitness distributions and these will be investigated in a companion paper [11]. From now on we assume that  $\mu$  is a probability measure supported by a bounded subinterval of the positive reals. Without loss of generality we assume that  $\mu$  has essential supremum equal to one. To avoid degeneracies we also assume that  $\mu$  has no atom at one.

We now describe our three main examples motivating our work.

**Example 1:** *Branching process with selection and mutation.*

This model is a stochastic *house-of-cards model* in a similar vein as Kingman's model (which is deterministic and much easier to analyse, see [18, 12]). We start with a single individual with a genetic fitness chosen according to  $\mu$ . Individuals never die and give birth to new individuals with a rate equal to their genetic fitness, the different reproduction rates causing the selection effect. When a new individual is born it is a mutant with probability  $\beta$ , in which case it gets a fitness drawn independently of everything else from  $\mu$ . If the new individual is not a mutant, it inherits the fitness of its parent. The model corresponds to the parameter choice  $\gamma = 1 - \beta$  in our framework. Observe that a mutation causes the complete loss of genetic information in the affected individual's ancestry, pictorially speaking 'the genetic house of cards collapses'. This is why the term house-of-cards model is used for this process, see [15] for a discussion of the relevance of these models in the theory of evolution.

The number of families  $M(t)$  corresponds to the number of mutants in the population at time  $t$ . We can describe the process  $(M(t))_{t \geq 0}$  as a Crump-Mode-Jagers process, using the framework of [20]. A mutant  $x$  born at time  $\tau$  with fitness  $f$  produces new mutants at ages according to a random point process  $\xi$ . This process is a Cox process, i.e. a Poisson process with a random intensity measure  $\beta f \phi_x(s) ds$ . The function  $\phi_x(s)$  is given by the size at time  $\tau + s$  of the family founded by the mutant and is therefore a Yule processes with intensity  $(1 - \beta)f$ .

The key assumption for the convergence theory of Crump-Mode-Jagers processes is the existence of a *Malthusian parameter*, i.e. an  $\alpha > 0$  such that

$$1 = \int_0^\infty e^{-\alpha s} \mathbb{E}\xi(ds).$$

In our case we have, for  $\alpha \geq 1 - \beta$ , that

$$\int_0^\infty e^{-\alpha s} \mathbb{E}\xi(ds) = \mathbb{E} \int_0^\infty e^{-\alpha s} \beta f \phi_x(s) ds = \beta \int f \int_0^\infty e^{-\alpha s + (1-\beta)fs} ds \mu(df) = \beta \int \frac{f}{\alpha - (1-\beta)f} \mu(df).$$

Hence a Malthusian parameter exists if and only if  $\frac{\beta}{1-\beta} \int \frac{f}{1-f} \mu(df) \geq 1$ . If this condition fails, the classical convergence theory of Crump-Mode-Jagers processes fails and very little is known about this case. In particular, in our model the precise asymptotics of  $M(t)$  is unknown. We show that in this case a phenomenon of condensation occurs, which loosely speaking means that a positive proportion of individuals have fitnesses converging to the maximal possible value. Key questions motivating this project are: *How fast is this convergence, when did the mutations arise that form the condensate, and how many mutations contribute to the condensate?*

**Example 2:** *Preferential attachment tree of Bianconi and Barabási.*

This model is originally a discrete time network model. Putting it into our framework means embedding it into continuous time, a technique heavily advocated by Janson [16], who attributes the method to Athreya and Karlin [1], and by Bhamidi [6]. The network is constructed successively, starting with one vertex which is formally given degree one. The vertex is given a fitness, randomly chosen according to  $\mu$ . At every time step a new vertex is introduced, equipped with a fitness, randomly chosen according to  $\mu$ , and linked to one of the existing vertices. The probability of an existing vertex being chosen is proportional to the product of its fitness and its degree at the time when the new vertex is introduced. As new vertices prefer to attach to existing vertices of high degree and high fitness, this is called a *preferential attachment* model.

In our representation we choose  $\beta = \gamma = 1$  and observe the system at the birth times of individuals. We think of every family as a vertex in the network, and of the size of a family as its degree. Note that when the  $n$ th birth event takes place, it arises in each of the existing families with a probability proportional to the product of its fitness and its degree. At the birth event a new family is founded, i.e. a new vertex is introduced, and simultaneously the family that has given birth is increased in size by one, meaning that the degree of the corresponding vertex is incremented by one. Our representation only keeps track of the vertices and their degrees, not of the actual edges. But this does not matter as the main object of interest for us is the long-term behaviour of the degree-weighted fitness distribution, which coincides with the empirical fitness distribution in our framework.

This model was analysed by Borgs et al. [7] who proved the existence of a *innovation-pays-off phase* in which a proportion of the mass in the degree-weighted fitness distribution condenses in the maximal fitness. This behaviour was already predicted in [4] who called this phase the winner-takes-all phase, a heavily misleading name as we shall see below. The result is reproved in our Theorem 2.1. Borgs et al. [7] state as an open problem ‘*to give an exact quantitative description of the innovation-pays-off phase. [...] How are the links distributed among the highest fitnesses present in the system at any given time? At what rate are new nodes with higher fitness taking over?*’ Our main aim here is to make progress on this problem.

**Example 3:** *Generalised Pólya urns.*

A class of generalised Pólya urns also falls into our framework, with general parameters  $\beta, \gamma > 0$  and  $\mu$  as above. It can be described as an urn containing balls of different colours. Every colour has a given *activity* chosen independently according to  $\mu$ . At time zero, the urn contains one ball of colour 1. At every time step, a ball is drawn at random from the urn with probability proportional to its activity. Then, the drawn ball is put back into the urn together with one or two new balls, at most one ball of the same and one of a new colour. A ball with the same colour is chosen with probability  $\gamma$ , and a ball of a new colour with probability  $\beta$ . New colours are chosen independently according to  $\mu$ . To embed the urn model into our framework we again look at the times of birth events. Observe that  $\Xi_t$  is now the empirical distribution of activities in the urn at time  $t$ .

Such generalised Pólya urns have apparently not been studied so far in full generality. Janson [16] is looking at the case where  $\mu$  is finitely supported, in which the condensation phenomenon, which is of interest to us, cannot arise. A related model has been studied by Chung et al. [8] who draw balls depending in a non-linear way on the distribution of colours in the urn, and by Collecchio et al. [9] who allow for a time-dependent replacement rule. Their main focus is on the question whether there can be an unbounded number of balls of more than one colour, and if not which colour eventually dominates. In our setup all colours will have an unbounded number of balls and we show that *the asymptotic proportion of balls of any colour goes to zero* uniformly as time goes to infinity.

## 2 Statement of the results

The reinforced branching process is described by the following family of random variables. We denote by

- $N(t)$  the total size of the population at time  $t$ ,
- $M(t)$  the number of different families at time  $t$ ,
- $\sigma_n$  the time of the  $n^{\text{th}}$  birth event,
- $\tau_n$  the time of the foundation of the  $n^{\text{th}}$  family,
- $Z_n(t)$  the size of the  $n^{\text{th}}$  family at time  $t$  (if  $n > M(t)$  we set  $Z_n(t) = 0$ ), and
- $F_n$  the fitness of the  $n^{\text{th}}$  family.

We are first interested in the empirical fitness distribution  $\Xi_t$  at time  $t$ , defined in (1). The asymptotic behaviour of this empirical distribution shows a phase transition between a fluid phase and a condensation phase. The condition for condensation is

$$\frac{\beta}{\beta + \gamma} \int_0^1 \frac{1}{1-x} d\mu(x) < 1 \quad \text{or, equivalently,} \quad \frac{\beta}{\gamma} \int_0^1 \frac{x}{1-x} d\mu(x) < 1, \quad (\text{Cond})$$

as stated in the following theorem.

### Theorem 2.1 (*Existence of a condensation phase*)

If (Cond) fails, then there exists a unique  $\lambda^* \in [\gamma, \beta + \gamma)$  such that

$$\frac{\beta}{\beta + \gamma} \int_0^1 \frac{\lambda^*}{\lambda^* - \gamma x} d\mu(x) = 1,$$

otherwise let  $\lambda^* := \gamma$ . In both cases

- the empirical mean fitness  $\int_0^1 x \Xi_t(dx)$  converges almost surely to  $\lambda^*/(\beta + \gamma)$ ,
- and there exists a probability measure  $\pi$  such that, almost surely, the empirical fitness distribution  $\Xi_t$  converges weakly to  $\pi$ .

The limit measure  $\pi$  of the empirical fitness distribution is given

(a) if (Cond) fails by

$$d\pi(x) = \frac{\beta}{\beta + \gamma} \frac{\lambda^*}{\lambda^* - \gamma x} d\mu(x).$$

(b) if (Cond) is true by

$$d\pi(x) = \frac{\beta}{\beta + \gamma} \frac{1}{1-x} d\mu(x) + \omega(\beta, \gamma) \delta_1(dx),$$

where

$$\omega(\beta, \gamma) := 1 - \frac{\beta}{\beta + \gamma} \int_0^1 \frac{1}{1-x} d\mu(x) > 0.$$

**Remark 1.** It is easy to see from the law of large numbers that

$$\frac{M(t)}{N(t)} \longrightarrow \frac{\beta}{\beta + \gamma} \quad \text{almost surely.}$$

Hence it is equivalent to ask for the absolute growth of either of the processes  $(M(t) : t > 0)$  or  $(N(t) : t > 0)$ . Given the population at time  $\sigma_n$  of the  $n$ th birth event, the waiting time  $\sigma_{n+1} - \sigma_n$  until the next individual is born is exponentially distributed with rate  $N(\sigma_n) \int x \Xi_{\sigma_n}(dx) \sim n\lambda^*$ , where we have used that  $N(\sigma_n) \sim (\beta + \gamma)n$  by the law of large numbers. Hence  $\sigma_n \sim \frac{1}{\lambda^*} \log n$  and, in particular, we obtain, almost surely,

$$\lim_{t \uparrow \infty} \frac{1}{t} \log N(t) = \lambda^*.$$

If there is no condensation we can improve this to convergence of  $e^{-t\lambda^*} N(t)$  to a positive random variable, using the arguments sketched in Section 3.2 below. But fine results about the growth of the population in the condensation phase are hard to obtain, see also Section 8.

**Remark 2.** We denote the part of the limit mass  $\pi$  which is absolutely continuous with respect to  $\mu$  as *bulk* and the part concentrated in the maximal fitness as *condensate*. The theorem shows that in the condensation phase, i.e. if (Cond) holds, we are seeing a phenomenon of *self-organised criticality*, as the number of individuals in the bulk and in the condensate are always kept on the same order of magnitude, without any tuning of parameters. In Dereich [10] one can see that for a model without self-organisation it can be rather complicated to tune the parameters in such a way that one has coexistence of bulk and condensate.

Our interest in this paper lies in the *emergence* of the condensate, i.e. how the condensate manifests itself at large finite times. Following the discussion of Bose-Einstein condensation in van den Berg et al. [3] two alternative scenarios are possible:

- For the largest family, the proportion of individuals belonging to this family in the overall population at time  $t$  is asymptotically positive. This phenomenon of *macroscopic occupancy* arises in condensation of the free Bose gas below a critical temperature, see [3].
- No individual family makes an asymptotically positive contribution. Instead, it is a collective effort of a growing number of families to form the condensate. This phenomenon is called *non-extensive condensation*. van den Berg et al. [3] have shown that this occurs in the free Bose gas for an intermediate temperature range.

We shall see in Theorem 2.4 that in our model under a natural assumption on  $\mu$  the second scenario prevails. To show this we need to investigate the behaviour of the largest family in our system. This requires some regularity assumptions on  $\mu$ . We henceforth assume that the fitness distribution  $\mu$  has a regularly varying tail in one, meaning that there exists  $\alpha > 1$  and a slowly varying function  $\ell$  with

$$\mu(1 - \varepsilon, 1) = \varepsilon^\alpha \ell(\varepsilon). \quad (\text{RV})$$

This corresponds to the most common type of behaviour of  $\mu$  at its tip that allows a condensation phase.

We start with a heuristic consideration. Suppose  $t > 0$  is given. At any time  $s \in (0, t)$  there are  $\exp((\lambda^* + o(1))s)$  families in the system and by an extreme value calculation the largest fitness in this number of families is of order  $1 - \exp(-(\frac{\lambda^*}{\alpha} + o(1))s)$ . Until time  $t > s$  the family achieving this fitness has time  $t - s$  to grow and therefore has size of order  $\exp(\gamma(t - s)(1 - e^{-\frac{\lambda^*}{\alpha}s}))$ . We therefore expect the birth time  $s$  of the maximal family at time  $t$  to be around the maximiser of this expression over  $0 < s < t$ . This maximum occurs roughly at time  $s \sim \alpha/\lambda^* \log t$ .

For the rigorous results we replace this time  $s$  by the stopping time

$$T(t) := \inf\{s \geq 0 : M(s) \geq n(t)\} \quad \text{where} \quad n(t) := \left\lceil \frac{1}{\mu(1-t^{-1}, 1)} \right\rceil,$$

which allows us precise control over the number of families in the system. Note that  $T(t) \sim \alpha/\lambda^* \log t$ , as can be seen by putting  $M(s) = \exp((\lambda^* + o(1))s)$  and  $\log n(t) = (\alpha + o(1)) \log t$ . Our heuristics suggests that the dominant families of the population at time  $t$  are born in a window around time  $T(t)$ , have fitness  $F_n$  with  $1 - F_n$  of order  $1/t$ , and size of order  $\exp(\gamma(t - T(t)))$ .

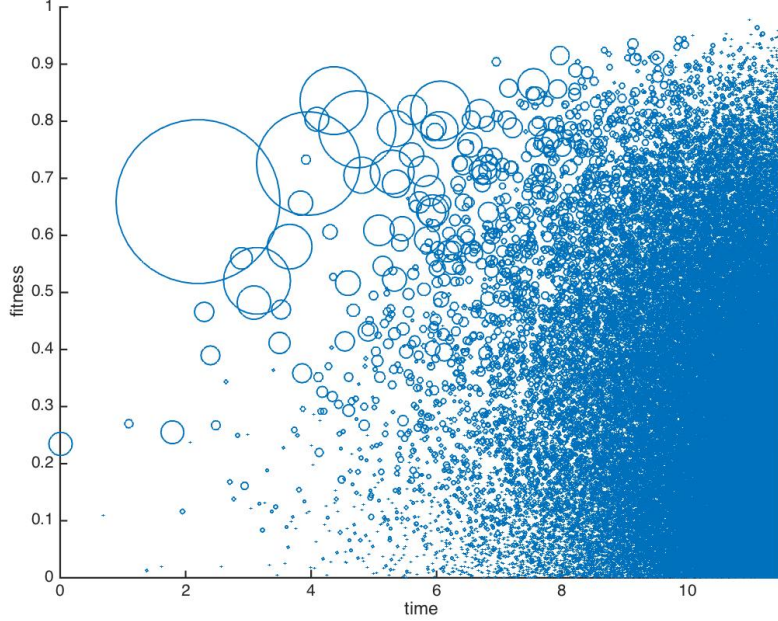


FIGURE 1 – A simulation of a reinforced branching process in the condensation case. Parameters are  $\mu(dx) = 3(1-x)^2 dx$  and  $\beta = \gamma = 1$ . Each family is represented by a circle with area proportional to its size at time  $t = 12$  and centred at its time of birth (horizontal axis) and its fitness (vertical axis). Simulation courtesy of Anna Senkevich.

To confirm this intuition we zoom into this window by considering the point process

$$\Gamma_t = \sum_{n=1}^{M(t)} \delta(\tau_n - T(t), (t - \tau_n)(1 - F_n), e^{-\gamma(t-T(t))} Z_n(t)),$$

where  $\delta(x)$  is the Dirac mass in  $x$ .

**Theorem 2.2 (Poisson limit)**

Under assumption (RV) the point process  $(\Gamma_t)_{t \geq 0}$  converges vaguely on the space  $[-\infty, \infty] \times [0, \infty] \times (0, \infty]$  to the Poisson point process with intensity measure

$$d\zeta(s, f, z) = \alpha f^{\alpha-1} \lambda^* e^{\lambda^* s} e^{-z e^{\gamma(s+f)}} e^{\gamma(s+f)} ds df dz.$$

**Remark 3.** Note the compactifications at  $\pm\infty$  in Theorem 2.2. As the limiting point process has a continuous density, Theorem 2.2 implies that all mass of  $\Gamma_t$  that asymptotically accumulates at infinity in one of the first two components, must escape at zero in the last component, meaning that the only way points can disappear in the limit is because the corresponding family size is small relative to the normalisation.

**Remark 4.** As there is no scaling in the first component of  $\Gamma_t$ , the limit theorem focuses on a time window of constant width around  $T(t)$ . The theorem shows that this is wide enough to capture the largest family at time  $t$ . However, it turns out that in the condensation phase this is *not* wide enough to capture *all* families that contribute to the condensate. This is why important questions on the emergence of the condensate remain open in this paper, see for example the first two problems in Section 8.

Our Poisson limit result, Theorem 2.2, readily implies the following distributional limits (denoted by  $\implies$ ) for the size, fitness and birth time of the largest family.

**Corollary 2.3** (Limits of family characteristics).

(i) Asymptotically, as  $t \rightarrow \infty$ ,

$$e^{-\gamma(t-T(t))} \max_{n \in \mathbb{N}} Z_n(t) \implies W^{-\frac{\gamma}{\lambda^*}},$$

where  $W$  is exponentially distributed with parameter  $\Gamma(\alpha + 1)\Gamma(1 + \frac{\lambda^*}{\gamma})(\lambda^*)^{-\alpha}$ .

(ii) Under (Cond), denoting by  $V(t)$  the fitness of the family of maximal size at time  $t$ , as  $t \rightarrow \infty$ , we have

$$t(1 - V(t)) \implies V,$$

where  $V$  is Gamma-distributed with scale parameter  $\lambda^*$  and shape parameter  $\alpha$ .

(iii) Denoting by  $S(t)$  the birth time of the family of maximal size at time  $t$ , as  $t \rightarrow \infty$ , we have

$$S(t) - T(t) \implies U,$$

where  $U$  is a real valued random variable.

**Remark 5.** The birth time of the family of maximal size at time  $t$  is of asymptotic order  $T(t) + O(1)$  and hence (as seen above) of leading order  $\alpha/\lambda^* \log t$ . This answers the question of Borgs et al. [7] about the rate at which new nodes with higher fitness become the leading influence in the population, see Figure 1 for a simulation.

**Theorem 2.4** (*The winner does not take it all*)

Under assumption (RV) the size of the largest family is negligible relative to the overall population size, i.e.

$$\lim_{t \rightarrow \infty} \frac{\max_{n \in \{1, \dots, M(t)\}} Z_n(t)}{N(t)} = 0, \text{ in probability.}$$

**Remark 6.** Theorem 2.4 means that asymptotically no single family contributes a positive proportion of the total mass, hence if there is condensation it is always *non-extensive*. This means in the context of Example 2 that no vertex attracts a positive fraction of the edges in the network. This is at odds with the informal description of condensation in the preferential attachment networks by Bianconi and Barabási [4], who are stating that ‘*the fittest node [is] acquiring a finite fraction of the links, independent of the size of the network.*’ It is also at odds with more recent work of Godrèche and Luck [14] who use a numerical study and further analysis based on it to conclude that asymptotically there is an unbounded number of macroscopic families. Apparently the phenomenon we investigate here is too subtle to be reliably captured by non-rigorous techniques. In the context of Example 3 our theorem states that the proportion of balls of any colour goes to zero, uniformly over all colours.

The remainder of this paper is organised as follows. In Section 3 we prove Theorem 2.1 by applying the theory of general branching processes. Section 4 contains an explicit construction of our model and uses this to give crude bounds on the rate of growth of the branching process. These will be used in Section 5 to derive a local version of Theorem 2.2, i.e. a version without the essential compactifications of the underlying space. Section 6 provides the estimates need to compactify the space, and in Section 7 we complete the proof of Theorem 2.2 and derive Corollary 2.3 and Theorem 2.4. The final section, Section 8, lists some interesting open problems.

### 3 Proof of the condensation phenomenon

In the last years a couple of techniques were developed to prove limit theorems for empirical distributions in networks and related structures, see for example Borgs et al. [7], Bhamidi [6] and Dereich and Ortgiese [13]. We now indicate how the theory of general branching processes can be used to prove Theorem 2.1. Our method is similar to the one described in [6] but circumvents the use of multitype branching.



### 3.1 The standard construction of the model

We start with a construction of our model on an explicit probability space. Let

- $F$  be a  $\mu$ -distributed random variable,
- given  $F$  let  $Y = (Y(t) : t \geq 0)$  be an independent Yule process with rate  $\gamma F$ ,
- given  $F, Y$  we define a simple point process  $\Pi = (\Pi(t) : t \geq 0)$  as  $\Pi = \Pi^{(1)} + \Pi^{(2)}$  where  $\Pi^{(1)}$  only jumps at the jumps of  $Y$ , and does so independently for every jump with probability  $\beta + \gamma - 1/\gamma$ , and  $\Pi^{(2)}$  is an independent, inhomogeneous Poisson process with intensity measure  $(1 - \gamma)FY(t) dt$ .

We let  $(\Omega, \mathcal{F}, \mathbb{P})$  be the countable product of the joint law of  $(F, Y, \Pi)$  and denote the coordinate process by  $(F_n, Y_n, \Pi_n)$ , for  $n \in \mathbb{N}$ . We let  $\tau_1 = 0$  and  $Z_1(t) = Y_1(t)$  and iteratively define, for  $n \in \{2, 3, \dots\}$ ,

$$\tau_n = \inf\{t > \tau_{n-1} : \exists m \in \{1, \dots, n-1\} \text{ with } \Delta \Pi_m(t - \tau_m) = 1\} \quad (2)$$

and

$$Z_n(t) = \begin{cases} Y_n(t - \tau_n), & \text{if } t \geq \tau_n \\ 0, & \text{otherwise.} \end{cases}$$

We let  $M(t) = \max\{n : \tau_n \leq t\}$ , set  $N(t) = \sum_{n=1}^{M(t)} Z_n(t)$ , and denote by  $\sigma_1, \sigma_2, \dots$  the jump times of  $(N(t) : t \geq 0)$ . It is obvious that this construction defines the reinforced branching process described in the introduction. Indeed  $(Y_n(t - \tau_n) : t \geq \tau_n)$  gives the times of birth of new individuals in the  $n$ th family, and  $(\Pi_n(t - \tau_n) : t \geq \tau_n)$  the times of creation of the new families which descend directly from the  $n$ th family.

For later reference we now recall some facts about Yule processes.

**Lemma 3.1.** *Let  $Y$  be a Yule process with rate  $\lambda$ . Then,*

- (a)  $(e^{-\lambda t} Y(t))_{t \geq 0}$  is a uniformly integrable martingale.
- (b) The almost sure limit  $\lim_{t \rightarrow \infty} e^{-\lambda t} Y(t)$  is standard exponentially distributed.
- (c) For  $u \in [0, 1)$  one has

$$\mathbb{E} \left[ \sup_{t \geq 0} \exp\{ue^{-\lambda t} Y(t)\} \right] \leq \frac{4}{1-u}. \quad (3)$$

- (d) Denote by  $T_n = \inf\{s \geq 0 : Y(s) \geq n\}$ . Then for every  $\varepsilon > 0$  with high probability as  $\kappa \rightarrow \infty$  for all  $n_0, n \geq \kappa$

$$\frac{1}{\lambda} \log \frac{n}{n_0} - \varepsilon \leq T_n - T_{n_0} \leq \frac{1}{\lambda} \log \frac{n}{n_0} + \varepsilon.$$

**Proof.** (a) and (b) are standard and proofs can be found in Athreya and Ney [2]. Denote the martingale limit in (b) by  $A$ . For the proof of (c) note that  $(\exp\{ue^{-t} Y(t)/2\} : t \geq 0)$  is a sub-martingale by Jensen's inequality. Doob's martingale inequality then gives

$$\mathbb{E} \left[ \sup_{t \geq 0} (\exp\{ue^{-t} Y(t)/2\})^2 \right] \leq 4 \mathbb{E}[\exp\{uA\}] = \frac{4}{1-u}.$$

To prove (d) we may assume, without loss of generality, that  $\lambda = 1$ . Consider the martingale given by  $\xi_t = e^{-t} Y(t)$ , and let  $R(\kappa) := \sup\{\frac{\xi_s}{\xi_u} : s, u \geq T_\kappa\}$ . By (b),  $(\xi_t)_{t \geq 0}$  has an almost surely finite, strictly positive limit and one has  $\lim_{\kappa \rightarrow \infty} R(\kappa) = 1$ , in probability. Further

$$\frac{Y(t + T_{n_0})}{Y(T_{n_0})} = e^t \frac{\xi_{t+T_{n_0}}}{\xi_{T_{n_0}}} \in \left[ \frac{1}{R(\kappa)} e^t, R(\kappa) e^t \right].$$

An application of the estimate for all  $n, n_0 \geq \kappa$  with  $t = T_n - T_{n_0}$  gives that

$$\frac{n}{n_0} \in \left[ \frac{1}{R(\kappa)} e^{T_n - T_{n_0}}, R(\kappa) e^{T_n - T_{n_0}} \right].$$

Taking logarithms and recalling that  $R(\kappa)$  tends to 1 yields the statement.  $\square$

### 3.2 General branching process theory

The processes  $(M(t): t > 0)$  is a general branching process, or Crump-Mode-Jagers process, with the laws of offspring times given by the point process  $(\Pi(t): t > 0)$ . Nerman [20] provides a strong law of large numbers for this class of processes under the assumption that there exists  $\lambda^* > \gamma$ , called the *Malthusian parameter*, such that

$$\int_0^\infty e^{-\lambda^* s} \mathbb{E}\Pi(ds) = 1.$$

An easy calculation (which we skip since it is already carried out in detail in the particular case of Example 1 above) shows that this is equivalent to

$$\frac{\beta}{\beta + \gamma} \int_0^1 \frac{\lambda^*}{\lambda^* - \gamma x} d\mu(x) = 1.$$

Suppose that  $\phi = \phi[F, Y, \Pi]: [0, \infty) \rightarrow \mathbb{N}_0$  is a cadlag process taking values in the nonnegative integers, such that  $\phi(t)$  is interpreted as a score assigned to a family  $t$  time units after its foundation. We assume the function  $t \mapsto \mathbb{E}[\phi(t)]$  is almost everywhere continuous and there exists  $h: [0, \infty) \rightarrow (0, \infty)$  integrable, bounded and non-increasing such that

$$\mathbb{E}\left[\sup_{t \geq 0} \frac{e^{-\lambda^* t} \phi(t)}{h(t)}\right] < \infty.$$

Letting  $\phi_n = \phi[F_n, Y_n, \Pi_n]$  we define the *score of the population* at time  $t$  as

$$Z^\phi(t) = \sum_{n: \tau_n < t} \phi_n(t - \tau_n).$$

We define

$$m_\infty^\phi = \frac{\int_0^\infty e^{-\lambda^* t} \mathbb{E}\phi(t) dt}{\int_0^\infty t e^{-\lambda^* t} \mathbb{E}\Pi(dt)}.$$

Nerman [20] shows that there exists a positive random variable  $W$ , not depending on  $\phi$ , such that

$$\lim_{t \uparrow \infty} e^{-\lambda^* t} Z^\phi(t) = W m_\infty^\phi \quad \text{almost surely.} \quad (4)$$

Under the assumption that (Cond) fails and  $\lambda^* > \gamma$  we can apply this result with  $\phi(t) = Y(t)$ , which satisfies the assumptions by Lemma 3.1, to get

$$\lim_{t \uparrow \infty} e^{-\lambda^* t} N(t) = W m_\infty^\phi \quad \text{almost surely.}$$

Choosing  $\phi(t) = Y(t) \mathbb{1}\{F \geq 1 - x\}$ , for  $0 < x < 1$ , and combining with the above gives

$$\lim_{t \uparrow \infty} \Xi_t[1 - x, 1] = \frac{\beta}{\beta + \gamma} \int_{1-x}^1 \frac{\lambda^*}{\lambda^* - \gamma x} d\mu(x) \quad \text{almost surely,}$$

as required.

### 3.3 A coupling technique

To extend results to the case when  $\lambda^* = \gamma$ , or when no Malthusian parameter is available, we use a coupling technique. We look at the reinforced branching process at the times  $(\sigma_n)_{n \geq 1}$  of the birth events and abbreviate  $\hat{\Xi}_n := \Xi_{\sigma_n}$ .

Fix  $\varepsilon > 0$ . We define a discrete-time branching process whose empirical fitness distribution  $\hat{\Xi}_n^{(\varepsilon)}$  has the property that, for all  $n \geq 0$ ,  $(\hat{\Xi}_n, \hat{\Xi}_n^{(\varepsilon)}) \in \mathcal{S}$ , where  $\mathcal{S}$  is the subset of the set of pairs of probability measures on  $[0, 1]$  defined by  $\mathcal{S} := \{(\nu, \mu): \nu([a, b]) \geq \mu([a, b]) \text{ for all } a, b \in [0, 1 - \varepsilon]\}$ . Let  $(U_n)_{n \geq 1}$  be a sequence of i.i.d. random variables uniformly distributed on  $[0, 1]$ .

At time zero, the new process contains one family of fitness  $F_1 \mathbb{1}\{F_1 < 1 - \varepsilon\} + \mathbb{1}\{F_1 \geq 1 - \varepsilon\}$  and thus  $(\hat{\Xi}_0, \hat{\Xi}_0^{(\varepsilon)}) \in \mathcal{S}$ . Assume now that,  $(\hat{\Xi}_n, \hat{\Xi}_n^{(\varepsilon)}) \in \mathcal{S}$ . We construct the new process at time  $n + 1$  as follows:

- if a new family of fitness  $f$  is born at time  $n + 1$  (in the original process), then we add in the (new) process a new family of fitness  $f\mathbf{1}\{f < 1 - \varepsilon\} + \mathbf{1}\{f \geq 1 - \varepsilon\}$  born at time  $n + 1$ ;
- if an individual of fitness larger than  $1 - \varepsilon$  is born at time  $n + 1$  in the original process, then we add a new individual of fitness 1 born at time  $n + 1$ ;
- if an individual of fitness  $f < 1 - \varepsilon$  is born at time  $n + 1$  in the original process, then if

$$U_{n+1} \leq \left( \frac{\widehat{\Xi}_n^{(\varepsilon)}(\{f\})}{\int_0^1 x d\widehat{\Xi}_n^{(\varepsilon)}(x)} \right) \left( \frac{\widehat{\Xi}_n(\{f\})}{\int_0^1 x d\widehat{\Xi}_n(x)} \right)^{-1},$$

we add an individual of fitness  $f$  born at time  $n + 1$ , otherwise, we add an individual of fitness 1.

By construction,  $(\widehat{\Xi}_{n+1}, \widehat{\Xi}_{n+1}^{(\varepsilon)}) \in \mathcal{S}$ . It is now easy to check that the new process is the discrete-time version of the reinforced branching process with fitness distribution  $\mu_\varepsilon := \mathbf{1}[0, 1 - \varepsilon)\mu + \mu[1 - \varepsilon, 1]\delta_1$ . Since

$$\frac{\beta}{\beta + \gamma} \int_0^1 \frac{d\mu_\varepsilon(x)}{1 - x} = \infty,$$

the new process admits a Malthusian parameter  $\lambda_\varepsilon$  and  $\lambda_\varepsilon \downarrow \gamma$  as  $\varepsilon \downarrow 0$ . We deduce that, for all  $0 \leq a, b < 1 - \varepsilon$ , we have

$$\lim_{n \rightarrow \infty} \widehat{\Xi}_n^{(\varepsilon)}([a, b]) = \lim_{t \rightarrow \infty} \widehat{\Xi}_t^{(\varepsilon)}([a, b]) = \frac{\beta}{\beta + \gamma} \int_a^b \frac{\lambda_\varepsilon}{\lambda_\varepsilon - \gamma x} d\mu(x)$$

almost surely. For all  $0 \leq a, b < 1$  and  $0 < \varepsilon < 1 - b$ , we thus have

$$\liminf_{t \rightarrow \infty} \Xi_t([a, b]) = \liminf_{n \rightarrow \infty} \widehat{\Xi}_n([a, b]) \geq \lim_{n \rightarrow \infty} \widehat{\Xi}_n^{(\varepsilon)}([a, b]) = \frac{\beta}{\beta + \gamma} \int_a^b \frac{\lambda_\varepsilon}{\lambda_\varepsilon - \gamma x} d\mu(x).$$

Letting  $\varepsilon \downarrow 0$  gives the lower bound. A similar argument gives a coupling with the reinforced branching process with  $\mu$  replaced by  $\mu^{(\varepsilon)} = \mathbf{1}[0, 1 - \varepsilon)\mu + \mu[1 - \varepsilon, 1]\delta_{1-\varepsilon}$ , and provides an upper bound, which is enough to conclude the proof of Theorem 2.1.

## 4 Estimates for the number of families in the population

The main difficulty in our model is that the time of birth of the  $n$ th family is not known with good accuracy. We now give a rough deterministic bound for the births occurring around the stopping times  $T(t)$ .

**Proposition 4.1.** *For all  $\varepsilon \in (0, \lambda^*)$ , we have with high probability as  $n_0 \rightarrow \infty$ , for all  $n \geq n_0$ ,*

$$\frac{1}{\lambda^* + \varepsilon} \log \frac{n}{n_0} - \varepsilon \leq \tau_n - \tau_{n_0} \leq \frac{1}{\lambda^* - \varepsilon} \log \frac{n}{n_0} + \varepsilon,$$

and, for all  $1 \leq n \leq n_0$ ,

$$\frac{1}{\lambda^* - \varepsilon} \log \frac{n}{n_0} - \varepsilon \leq \tau_n - \tau_{n_0} \leq \frac{1}{\lambda^* + \varepsilon} \log \frac{n}{n_0} + \varepsilon.$$

**Proof.** Recall that, when  $t$  goes to infinity,  $M(t) = e^{(\lambda^* + o(1))t}$ , implying that

$$\tau_{n_0} = \frac{1 + o(1)}{\lambda^*} \log n_0$$

which implies the estimate for arbitrarily fixed  $n$  as  $n_0$  goes to infinity. It thus suffices to prove the statement for all  $n, n_0 \geq \kappa$  with high probability as  $\kappa$  goes to infinity. Fix  $\varepsilon \in (0, \lambda^*)$  and  $\kappa \in \mathbb{N}$  and set  $\lambda_\pm^* := \lambda^* \pm \varepsilon$ . The stochastic process  $(M(s))_{s \geq 0}$  is a pure birth process with continuous compensator

$$\gamma(s) = \int_0^s \beta N(u) \int_0^1 x \Xi_u(dx) du.$$

We consider the stopping time  $S$  given by

$$S = \inf \left\{ s \geq \tau_\kappa : \frac{\beta N(s)}{M(s)} \int_0^1 x \Xi_s(dx) \notin [\lambda_-^*, \lambda_+^*] \right\},$$

and note that, by Theorem 2.1,  $S$  is infinite with high probability as  $\kappa \rightarrow \infty$ .

Let  $w_n := \tau_{n+1} - \tau_n$  be the inter-arrival times of  $(M(s))_{s \geq 0}$ . Observe that

$$\gamma(\tau_{n+1}) - \gamma(\tau_n) = \int_{\tau_n}^{\tau_{n+1}} \beta N(s) \int_0^1 x \Xi_s(dx) ds \in [\lambda_-^*(\tau_{n+1} - \tau_n)n, \lambda_+^*(\tau_{n+1} - \tau_n)n]$$

provided that  $\tau_{n+1} \leq S$ . Defining

$$w_n^\pm := \frac{\gamma(\tau_{n+1}) - \gamma(\tau_n)}{\lambda_\pm^* n},$$

we infer that  $w_n^+ \leq w_n \leq w_n^-$  for all  $n$  such that  $\tau_{n+1} \leq S$ , and the sequences  $(w_n^\pm)_{n \in \mathbb{N}}$  consist of independent exponentials with respective parameter  $\lambda_\pm^* n$  which are the inter-arrival times of Yule processes  $(Y_\pm(s))$  of respective rate  $\lambda_\pm^*$ . By Lemma 3.1, with high probability as  $\kappa \rightarrow \infty$ , for all  $n, n_0 \geq \kappa$ ,

$$\frac{1}{\lambda_\pm^*} \log \frac{n}{n_0} - \varepsilon \leq T_n^\pm - T_{n_0}^\pm \leq \frac{1}{\lambda_\pm^*} \log \frac{n}{n_0} + \varepsilon,$$

where  $(T_n^\pm)_{n \in \mathbb{N}}$  denotes the ordered sequence of jump times of  $(Y_\pm(s))_{s \geq 0}$ . Hence, for all  $n \geq n_0 \geq \kappa$ ,

$$\tau_n - \tau_{n_0} = \sum_{k=n_0}^{n-1} w_k \leq \sum_{k=n_0}^{n-1} w_k^- = T_n^- - T_{n_0}^- \leq \frac{1}{\lambda_-^*} \log \frac{n}{n_0} + \varepsilon.$$

Similarly, it follows that with high probability, for  $\kappa \leq n \leq n_0$ ,

$$\tau_n - \tau_{n_0} \leq T_n^+ - T_{n_0}^+ \leq \frac{1}{\lambda_+^*} \log \frac{n}{n_0} + \varepsilon.$$

The converse bound follows in complete analogy. □

## 5 Local convergence

The aim of this section is to prove the following proposition.

**Proposition 5.1.** *Under assumption (RV) one has convergence in distribution of the point processes*

$$\Gamma_t = \sum_{n=1}^{M(t)} \delta_{(\tau_n - T(t), (t - \tau_n)(1 - F_n), e^{-\gamma(t - T(t))} Z_n(t))}$$

*to the Poisson point process with intensity*

$$d\zeta(s, f, z) = \alpha f^{\alpha-1} \lambda^* e^{\lambda^* s} e^{-ze^{\gamma(f+s)}} e^{\gamma(f+s)} ds df dz,$$

*where we endow the set of locally finite measures on  $(-\infty, \infty) \times [0, \infty) \times [0, \infty]$  with the topology of vague convergence.*

Proposition 5.1 is a straightforward consequence of the following result.

**Proposition 5.2.** *Under assumption (RV) we have vague convergence of the point process*

$$\Psi_t = \sum_{n=1}^{M(t)} \delta_{(\tau_n - T(t), (t - \tau_n)(1 - F_n), e^{-\gamma F_n(t - \tau_n)} Z_n(t))}$$

*to the Poisson point process with intensity*

$$d\zeta^*(s, f, z) = \alpha f^{\alpha-1} \lambda^* e^{\lambda^* s} e^{-z} ds df dz$$

*on  $(-\infty, \infty) \times [0, \infty) \times [0, \infty]$ .*

**Proof of Proposition 5.1.** This follows directly from the fact that the point process  $\Gamma_t$  is the image of  $\Psi_t$  by the continuous function  $(s, f, z) \mapsto (s, f, e^{-\gamma(s+f)z})$ , and that  $\zeta$  is the image of  $\zeta^*$  by the same continuous function.  $\square$

The proof of Proposition 5.2 consists of the following two steps. We approximate  $\Psi_t$  by a point process

$$\Psi_t^* = \sum_{n \in \mathbb{N}} \delta_{(\tau_n^*(t) - T(t), t(1 - F_n), \xi_n)},$$

where the birth times  $\tau_n$  are replaced by approximate birth times

$$\tau_n^*(t) = T(t) + \frac{\log n/n(t)}{\lambda^*}$$

and the rescaled family sizes  $e^{-\gamma F_n(t - \tau_n)} Z_n(t)$  by their limits

$$\xi_n = \lim_{u \rightarrow \infty} e^{-\gamma F_n(u - \tau_n)} Z_n(u).$$

In the approximating process  $\Psi_t^*$  the components are decoupled, which makes it easier to study. We prove that

- (1) this approximation process converges vaguely to the Poisson point process of intensity  $\zeta^*$ , and
- (2)  $\Psi_t^*$  is close enough to  $\Psi_t$  to imply Proposition 5.2.

The two steps correspond to the two lemmas below.

**Lemma 5.3.**  $(\Psi_t^*)_{t \geq 0}$  converges vaguely on  $[-\infty, \infty) \times [0, \infty) \times [0, \infty]$  to the Poisson point process with intensity  $\zeta^*$ .

**Proof.** We apply Kallenberg's theorem, see [22, Proposition 3.22]. Since  $\zeta^*$  is diffuse, to prove Lemma 5.3, it is enough to show that, for every precompact relatively open box  $B \subset [-\infty, \infty) \times [0, \infty) \times [0, \infty]$ ,

- (a)  $\mathbb{P}(\Psi_t^*(B) = 0) \rightarrow \exp(-\zeta^*(B))$ , as  $t \uparrow \infty$ , and
- (b)  $\mathbb{E}\Psi_t^*(B) \rightarrow \zeta^*(B)$ , as  $t \uparrow \infty$ .

It suffices to consider nonempty boxes  $B$  of the form  $(a_0, a_1) \times (b_0, b_1) \times (c_0, c_1)$  since, almost surely, neither the point processes  $\Gamma_t$  nor the limiting Poisson process put points on the boundary  $\partial([-\infty, \infty) \times [0, \infty) \times [0, \infty])$ . Here  $a_0 = -\infty$  and  $c_1 = \infty$  is an allowed choice. Note that

$$\zeta^*(B) = (e^{\lambda^* a_1} - e^{\lambda^* a_0})(b_1^\alpha - b_0^\alpha)(e^{-c_0} - e^{-c_1}).$$

- (a) By the construction of the probability space at the beginning of Section 4,  $(F_n, \xi_n)_{n \geq 1}$  is a sequence of i.i.d. random variables with each  $F_n$  being independent of  $\xi_n$ . Hence,

$$\begin{aligned} \mathbb{P}(\Psi_t^*(B) = 0) &= \prod_{n(t)e^{\lambda^* a_0} < n < n(t)e^{\lambda^* a_1}} \mathbb{P}(t(1 - F_n) \notin (b_0, b_1) \text{ or } \xi_n \notin (c_0, c_1)) \\ &= (1 - \mathbb{P}(t(1 - F_1) \in (b_0, b_1)) \mathbb{P}(\xi_1 \in (c_0, c_1)))^{r_{a_0, a_1}(t)}, \end{aligned}$$

where  $r_{a_0, a_1}(t)$  denotes the number of elements  $n \in \mathbb{N}$  with  $n(t)e^{\lambda^* a_0} < n < n(t)e^{\lambda^* a_1}$ .

We note that, as  $t \rightarrow \infty$ , we have  $r_{a_0, a_1}(t) \sim (e^{\lambda^* a_1} - e^{\lambda^* a_0})/\mu(1 - t^{-1}, 1)$  and, in view of Assumption RV,

$$\frac{\mu((1 - b_1/t, 1 - b_0/t))}{\mu(1 - 1/t, 1)} \sim b_1^\alpha - b_0^\alpha \quad \text{as } t \uparrow \infty. \quad (5)$$

Further  $\mathbb{P}(\xi_1 \in (c_0, c_1)) = e^{-c_0} - e^{-c_1}$ . Thus, as  $t \rightarrow \infty$ ,

$$\begin{aligned} \mathbb{P}(\Psi_t^*(B) = 0) &= (1 - \mu(1 - b_1/t, 1 - b_0/t) (e^{-c_0} - e^{-c_1}))^{r_{a_0, a_1}(t)} \\ &\sim \exp\left(-(e^{\lambda^* a_1} - e^{\lambda^* a_0}) \frac{\mu((1 - b_1/t, 1 - b_0/t))}{\mu(1 - 1/t, 1)} (e^{-c_0} - e^{-c_1})\right) \\ &\rightarrow \exp(-\zeta^*(B)). \end{aligned}$$

(b) To compute the limit of  $\mathbb{E}[\Psi_t^*(B)]$  we apply the asymptotic estimates from above,

$$\begin{aligned}\mathbb{E}[\Psi_t^*(B)] &= \sum_{n(t)e^{\lambda^* a_0} < n < n(t)e^{\lambda^* a_1}} \mu([1 - b_1/t, 1 - b_0/t]) \mathbb{P}(\xi_n \in [c_1, c_2]) \\ &= r_{a_0, a_1}(t) \mu([1 - b_1/t, 1 - b_0/t]) \mathbb{P}(\xi_1 \in [c_1, c_2]) \rightarrow \zeta^*(B).\end{aligned}$$

□

**Lemma 5.4.** *For all Lipschitz continuous, compactly supported functions  $f : (-\infty, \infty) \times [0, \infty) \times [0, \infty] \rightarrow \mathbb{R}$ ,*

$$\left| \int f d\Psi_t^* - \int f d\Psi_t \right| \rightarrow 0 \text{ in probability, as } t \uparrow \infty.$$

**Proof.** Let  $f$  be a Lipschitz continuous function supported on  $K = [-a, a] \times [0, b] \times [0, \infty]$  for  $a, b \geq 1$ . We have

$$\begin{aligned}& \left| \int f d\Psi_t - \int f d\Psi_t^* \right| \\ & \leq \sum_{n=1}^{M(t)} \left| f\left(\tau_n - T(t), (t - \tau_n)(1 - F_n), e^{-\gamma F_n(t - \tau_n)} Z_n(t)\right) - f\left(\tau_n^*(t) - T(t), t(1 - F_n), \xi_n\right) \right| \\ & \leq c \sum_{n \in I(t)} (|\tau_n - \tau_n^*(t)| + \tau_n(1 - F_n) + |e^{-\gamma F_n(t - \tau_n)} Z_n(t) - \xi_n|),\end{aligned} \quad (6)$$

where  $c$  is the Lipschitz constant of the function  $f$  and  $I(t)$  is the random set of indices  $n \in \mathbb{N}$  such that

$$|\tau_n - T(t)| \leq a \quad \text{and} \quad (t - \tau_n)(1 - F_n) \leq b \quad (a)$$

or

$$|\tau_n^*(t) - T(t)| \leq a \quad \text{and} \quad t(1 - F_n) \leq b. \quad (b)$$

For  $\varepsilon \in (0, 1/2)$  we denote by  $\mathcal{T}_\varepsilon(t)$  the event that the following properties hold,

- $|\tau_n - \tau_n^*(t)| \leq \varepsilon(1 + |\tau_n^*(t) - T(t)|)$  for all  $n \in \mathbb{N}$  and
- $T(t) \leq t/3$ .

We note that, in view of Proposition 4.1,  $\mathcal{T}_\varepsilon(t)$  holds with high probability as  $t \rightarrow \infty$ .

Let now  $\bar{I}(t) := \{n \in \mathbb{N} : |\tau_n^*(t) - T(t)| \leq 2a, t(1 - F_n) \leq 2b\}$ . We show that for  $t \geq 6a$  one has  $I(t) \subset \bar{I}(t)$  on  $\mathcal{T}_\varepsilon(t)$ . Suppose that  $n \in I(t)$  and that  $\mathcal{T}_\varepsilon(t)$  holds. It suffices to consider the case where condition (a) is satisfied. Then  $b \geq (t - \tau_n)(1 - F_n) \geq (t - T(t) - a)(1 - F_n) \geq t(1 - F_n)/2$ , which proves that the second inequality in the definition of  $\bar{I}(t)$  is satisfied. Let us further assume that  $n \leq n(t)$ . Then  $a \geq |\tau_n - T(t)| \geq (1 - \varepsilon)|\tau_n^*(t) - T(t)| - \varepsilon$ , which implies that

$$|\tau_n - \tau_n^*(t)| \leq \varepsilon(1 + |\tau_n^*(t) - T(t)|) \leq \varepsilon(1 + \frac{a + \varepsilon}{1 - \varepsilon}) \leq 4a\varepsilon.$$

The same inequality holds if  $n \geq n(t)$  and thus we have proved that  $I(t) \subset \bar{I}(t)$  on  $\mathcal{T}_\varepsilon(t)$  for all  $t \geq 6a$ .

We now consider the sum on the right hand side of (6), but taken over all  $n \in \bar{I}(t)$ . First note that, for  $n \in \bar{I}(t)$ , on  $\mathcal{T}_\varepsilon(t)$ , we have  $\tau_n \leq T(t) + (1 + \varepsilon)(\tau_n^*(t) - T(t)) + \varepsilon \leq t/3 + 3a$ , if  $n \geq n(t)$ , and  $\tau_n(1 - F_n) \leq 2b/t(T(t) + 3a)$ . Second we let, for  $n \in \mathbb{N}$  and  $s \geq 0$ ,

$$R_n(s) := \sup_{u \geq s} |e^{-u} Z_n(\tau_n + u/F_n) - \xi_n|$$

and using that for  $t \geq 4b$  and  $n \in \bar{I}(t)$  one has  $F_n \geq 1/2$  we conclude that for all  $t \geq \max(4b, 18a)$ ,

$$|e^{-\gamma F_n(t - \tau_n)} Z_n(t) - \xi_n| \leq R_n(\gamma F_n(t - \tau_n)) \leq R_n(\gamma t/4).$$

Hence we get that, for sufficiently large  $t$ , on  $\mathcal{T}_\varepsilon(t)$ ,

$$\begin{aligned}
& \left| \int f d\Psi_t - \int f d\Psi_t^* \right| \\
& \leq c \sum_{n \in \bar{I}(t)} (|\tau_n - \tau_n^*(t)| + \tau_n(1 - F_n) + |e^{-\gamma F_n(t - \tau_n)} Z_n(t) - \xi_n|) \\
& \leq c \sum_{n \in \bar{I}(t)} \left( 4a\varepsilon + \frac{2b}{t}(T(t) + 3a) + R_n(\gamma t/4) \right) \\
& \leq c |\bar{I}(t)| \left( 4a\varepsilon + \frac{2b}{t}(T(t) + 3a) \right) + c \sum_{n \in \bar{I}(t)} R_n(\gamma t/4).
\end{aligned}$$

By construction, the random processes  $(R_n)_{n \geq 1}$  are independent of  $(F_n)_{n \geq 1}$  and thus also of the random set  $\bar{I}(t)$ . We recall that, by Proposition 5.2,  $|\bar{I}(t)|$  converges in distribution to a Poisson distribution and  $\lim_{s \rightarrow \infty} R_n(s) = 0$ , almost surely. Hence,

$$\lim_{t \rightarrow \infty} \sum_{n \in \bar{I}(t)} R_n(\gamma st/4) = 0, \text{ in probability.}$$

Since further  $\lim_{t \rightarrow \infty} T(t)/t = 0$ , almost surely, we conclude that, with high probability,

$$\left| \int f d\Psi_t - \int f d\Psi_t^* \right| \leq 8\varepsilon ac(|\bar{I}(t)| + 1).$$

Recalling again that  $|\bar{I}(t)|$  converges in distribution and that  $\varepsilon \in (0, 1/2)$  can be made arbitrarily small we obtain convergence to zero in probability, as  $t \uparrow \infty$ .  $\square$

**Proof of Proposition 5.2.** Let  $f: \mathbb{R} \times [0, \infty) \times [0, \infty) \rightarrow \mathbb{R}$  be Lipschitz continuous and compactly supported. Combining Lemmas 5.3 and 5.4, together with Slutsky's lemma we get the desired result,

$$\int f d\Psi_t \Rightarrow \int f d\text{PPP}(\zeta^*) \quad \text{as } t \uparrow \infty,$$

where  $\text{PPP}(\zeta^*)$  denotes the Poisson point process with intensity  $\zeta^*$ .  $\square$

## 6 Negligibility of families outside the main window

To deduce Theorem 2.2 from Proposition 5.1, one has to control the contribution of the point process near the closed boundaries of  $[-\infty, +\infty] \times [0, +\infty] \times (0, +\infty]$ . We prove that the families that are born too late, or that are not fit enough, are too small to contribute in the limit. They *get absorbed by the open lower bound of the third coordinate*. We first provide a simple calculation, which is at the heart of our proofs. Recall that

$$n(t) = \left\lceil \frac{1}{\mu(1 - 1/t, 1)} \right\rceil.$$

**Lemma 6.1.** *Let  $F$  be a random variable with law  $\mu$ . There exists  $t_0 > 0$  such that, for all  $C \geq 0$ ,  $D > 0$ , there exists  $K = K(C, D) > 0$  such that*

$$\mathbb{E} \left[ \mathbf{1}\{F \leq 1 - \frac{C}{t}\} e^{-D(1-F)t} \right] \leq \frac{K}{n(t)} \text{ for all } t \geq t_0.$$

Moreover, for all  $D$ , we have  $\lim_{C \uparrow \infty} K(C, D) = 0$ .

To prove this lemma, we need Potter's bound, see Theorem 1.5.6 (ii) in [5]. Since  $\mu$  verifies Equation (RV), and is bounded from zero and infinity on every compact set of  $(0, 1]$ , for all  $\delta > 0$ , there exists a constant  $\Lambda = \Lambda(\delta)$  such that, for all  $0 < x, y \leq 1$ ,

$$\frac{\mu(1 - y, 1)}{\mu(1 - x, 1)} \leq \Lambda \left( \frac{y}{x} \right)^{\alpha + \delta}. \quad (7)$$

**Proof of Lemma 6.1.** Fix  $\delta > 0$ . We have

$$\begin{aligned}\mathbb{E} \left[ \mathbf{1}\{F \leq 1 - \frac{C}{t}\} e^{-D(1-F)t} \right] &= \int_0^\infty \mathbb{P} \left( \mathbf{1}\{F \leq 1 - \frac{C}{t}\} e^{-D(1-F)t} \geq x \right) dx = \int_0^\infty \mathbb{P} \left( 1 - \frac{\log 1/x}{Dt} \leq F \leq 1 - \frac{C}{t} \right) dx \\ &\leq \int_0^{e^{-CD}} \mu \left( 1 - \frac{\log 1/x}{Dt}, 1 \right) dx \leq \Lambda(\delta) \mu(1 - 1/t, 1) \int_0^{e^{-CD}} \left( \frac{1}{D} \log \frac{1}{x} \right)^{\alpha+\delta} dx,\end{aligned}$$

where  $\Lambda(\delta)$  is defined by the Potter's bound (see Equation (7)). Changing variables  $y = \frac{1}{D} \log \frac{1}{x}$  we get

$$\mathbb{E} \left[ \mathbf{1}\{F \leq 1 - \frac{C}{t}\} e^{-D(1-F)t} \right] \leq D\Lambda(\delta) \mu(1 - 1/t, 1) \int_C^\infty y^{\alpha+\delta} e^{-Dy} dy.$$

Recall that  $n(t) = \lfloor 1/\mu(1-1/t, 1) \rfloor$ , and let  $K(C, D) := D\Lambda(\delta) \int_C^\infty y^{\alpha+\delta} e^{-Dy} dy$  to conclude the proof.  $\square$

## 6.1 Contribution of the unfit families

**Lemma 6.2.** For every  $\eta > 0$  and  $c > 0$  there exists  $\kappa > 0$  such that, for all sufficiently large  $t$ , we have

$$\mathbb{P} \left( \max_{n \leq M(t)} \mathbf{1}\{F_n \leq 1 - \kappa/t\} Z_n(t) \geq c e^{\gamma(t-T(t))} \right) \leq \eta.$$

**Proof.** Let  $c > 0$  and  $\kappa > 0$ . We analyse the event that there exists a family with fitness  $F_n \leq 1 - \kappa/t$  and size  $e^{-\gamma(t-T(t))} Z_n(t) \geq c$ . To this end, we define the time-shifted version  $(Z_n^*(t) : t \in \mathbb{R})$  of the size of the  $n^{\text{th}}$  family by

$$Z_n^*(t) := Z_n(t + \tau_n - \tau_n^*),$$

where

$$\tau_n^* := \begin{cases} T(t) + \frac{1}{\lambda^* - \varepsilon} \log \frac{n}{n(t)} - \varepsilon & \text{if } n \leq n(t), \\ T(t) + \frac{1}{\lambda^* + \varepsilon} \log \frac{n}{n(t)} - \varepsilon & \text{if } n \geq n(t). \end{cases}$$

In view of Proposition 4.1, we have, with high probability as  $t \rightarrow \infty$ , that  $Z_n^*(t) \geq Z_n(t)$  for all  $n \geq 0$ . Recalling the construction of the probability space at the beginning of Section 4, we note that the family  $(A_n)_{n \geq 1}$  given by

$$A_n = \max_{s \geq \tau_n^*} \frac{Z_n^*(s)}{e^{\gamma F_n(s - \tau_n^*)}} = \max_{s \geq \tau_n} \frac{Z_n(s)}{e^{\gamma F_n(s - \tau_n)}}$$

forms a sequence of i.i.d. random variables which is independent of  $(F_n)_{n \geq 1}$ . Further,

$$\begin{aligned}\{Z_n(t) \geq c e^{\gamma(t-T(t))}\} &\subset \{Z_n^*(t) \geq c e^{\gamma(t-T(t))}\} \\ &\subset \{A_n \geq c e^{\gamma[(t-T(t)) - F_n(t - \tau_n^*)]}\} = \{A_n \geq c e^{\gamma[(1-F_n)(t-T(t)) - F_n(T(t) - \tau_n^*)]}\}.\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{P} \left( \max_{n \leq M(t)} \mathbf{1}\{F_n \leq 1 - \kappa/t\} Z_n(t) \geq c e^{\gamma(t-T(t))} \right) \\ \leq \mathbb{P}(T(t) \geq t/2) + \sum_{n=1}^\infty \mathbb{P}(A_n \mathbf{1}\{F_n \leq 1 - \kappa/t\} \geq c e^{\gamma[(1-F_n)t/2 - F_n(T(t) - \tau_n^*)]}) .\end{aligned} \tag{8}$$

In terms of  $\varphi(u) := \mathbb{P}(A_1 \geq u)$  we have

$$\mathbb{P} \left( A_n \mathbf{1}\{F_n \leq 1 - \kappa/t\} \geq c e^{\gamma[(1-F_n)t/2 - F_n(T(t) - \tau_n^*)]} \right) = \mathbb{E} \left[ \mathbf{1}\{F_n \leq 1 - \kappa/t\} \varphi \left( c e^{\gamma[(1-F_n)t/2 - F_n(T(t) - \tau_n^*)]} \right) \right].$$

By Lemma 3.1, we have  $\varphi(u) \leq C_0 e^{-u/2}$  so that

$$\begin{aligned}\mathbb{E} \left[ \mathbf{1}\{F_n \leq 1 - \kappa/t\} \varphi \left( c e^{\gamma[(1-F_n)t/2 - F_n(T(t) - \tau_n^*)]} \right) \right] \\ \leq C_0 \mathbb{E} \left[ \mathbf{1}\{F_n \leq 1 - \kappa/t\} \exp \left\{ -\frac{c}{2} e^{\gamma[(1-F_n)t/2 - F_n(T(t) - \tau_n^*)]} \right\} \right].\end{aligned}$$



Noting that  $T(t) - \tau_n^*$  is deterministic we get that the expectation on the right equals

$$\mathbb{E}\left[\mathbb{1}\{F \leq 1 - \kappa/t\} \exp\left\{-\frac{\varepsilon}{2}e^{\gamma[(1-F)^{t/2}-F(T(t)-\tau_n^*)]}\right\}\right],$$

where  $F$  is a random variable of law  $\mu$ . We now fix small numbers  $\delta, \varrho > 0$  and note that there exists a constant  $C_\varrho$  such that  $e^{-y} \leq C_\varrho y^{-\varrho}$  for all  $y \geq 0$ . Using this, and recalling the definition of  $\tau_n^*$ , we get, for  $n \geq n(t)$ ,

$$\begin{aligned} & \mathbb{E}\left[\mathbb{1}\{\delta \leq F \leq 1 - \kappa/t\} \exp\left\{-\frac{\varepsilon}{2}e^{\gamma[(1-F)^{t/2}-F(T(t)-\tau_n^*)]}\right\}\right] \\ & \leq \mathbb{E}\left[\mathbb{1}\{F \leq 1 - \kappa/t\} \exp\left\{-\frac{\varepsilon}{2}\left(\frac{n}{n(t)}\right)^{\frac{\delta\gamma}{\lambda^*+\varepsilon}}e^{\gamma[(1-F)^{t/2}-\varepsilon]}\right\}\right] \\ & \leq C_\varrho \left(\frac{\varepsilon}{2}\right)^{-\varrho} e^{\gamma\varrho\varepsilon} \left(\frac{n(t)}{n}\right)^{\varrho\frac{\gamma\delta}{\lambda^*+\varepsilon}} \mathbb{E}\left[\mathbb{1}\{F \leq 1 - \kappa/t\}e^{-\varrho\gamma(1-F)^{t/2}}\right]. \end{aligned}$$

We now apply Lemma 6.1 and get, for  $n \geq n(t)$ ,

$$\mathbb{E}\left[\mathbb{1}\{\delta \leq F \leq 1 - \kappa/t\} \exp\left\{-\frac{\varepsilon}{2}e^{\gamma[(1-F)^{t/2}-F(T(t)-\tau_n^*)]}\right\}\right] \leq C_\varrho \left(\frac{\varepsilon}{2}\right)^{-\varrho} e^{\gamma\varrho\varepsilon} \left(\frac{n(t)}{n}\right)^{\varrho\frac{\gamma\delta}{\lambda^*+\varepsilon}} \frac{K(\kappa, \varrho\gamma/2)}{n(t)}. \quad (9)$$

Similarly we get, for  $n \leq n(t)$ ,

$$\mathbb{E}\left[\mathbb{1}\{\delta \leq F \leq 1 - \kappa/t\} \exp\left\{-\frac{\varepsilon}{2}e^{\gamma[(1-F)^{t/2}-F(T(t)-\tau_n^*)]}\right\}\right] \leq C_\varrho \left(\frac{\varepsilon}{2}\right)^{-\varrho} e^{\gamma\varrho\varepsilon} \left(\frac{n(t)}{n}\right)^{\varrho\frac{\gamma}{\lambda^*-\varepsilon}} \frac{K(\kappa, \varrho\gamma/2)}{n(t)}. \quad (10)$$

Applying (9) with  $\varrho_+ > \frac{\lambda^*+\varepsilon}{\gamma\delta}$ , if  $n > n(t)$ , and (10) with  $\varrho_- < \frac{\lambda^*-\varepsilon}{\gamma}$ , if  $n \leq n(t)$ , we get

$$\begin{aligned} & \sum_{n=1}^{\infty} \mathbb{P}(A_n \mathbb{1}\{F_n \leq 1 - \kappa/t\} \geq c e^{\gamma[(1-F_n)^{t/2}-F_n(T(t)-\tau_n^*)]}) \\ & \leq C_0 C_{\varrho_-} \left(\frac{\varepsilon}{2}\right)^{-\varrho_-} e^{\gamma\varrho_- \varepsilon} \frac{K(\kappa, \varrho_- \gamma/2)}{n(t)} \sum_{n=1}^{n(t)} \left(\frac{n(t)}{n}\right)^{\varrho_- \frac{\gamma}{\lambda^*-\varepsilon}} \\ & \quad + C_0 C_{\varrho_+} \left(\frac{\varepsilon}{2}\right)^{-\varrho_+} e^{\gamma\varrho_+ \varepsilon} \frac{K(\kappa, \varrho_+ \gamma/2)}{n(t)} \sum_{n=n(t)+1}^{\infty} \left(\frac{n(t)}{n}\right)^{\varrho_+ \frac{\gamma\delta}{\lambda^*+\varepsilon}} + C_0 \mathbb{P}(F < \delta) \\ & \leq C (K(\kappa, \varrho_- \gamma/2) + K(\kappa, \varrho_+ \gamma/2)) + C_0 \mathbb{P}(F < \delta), \end{aligned}$$

where  $C$  is a constant not depending on  $\kappa$  or  $t$ , using that both sums are bounded by a constant multiple of  $n(t)$ . Recall that  $\lim_{\kappa \rightarrow \infty} K(\kappa, \varrho \pm \gamma/2) = 0$  and  $\mathbb{P}(F < \delta) \rightarrow 0$  as  $\delta \downarrow 0$ . Recalling (8) and noting that  $\mathbb{P}(T(t) \geq t/2) \rightarrow 0$ , as  $t \uparrow \infty$ , completes the proof.  $\square$

## 6.2 Contribution of the families born late

**Lemma 6.3.** *For every  $\eta > 0$  and  $c > 0$  there exists  $v > 1$  such that, for all sufficiently large  $t$ , we have*

$$\mathbb{P}\left(\max_{vn(t) \leq n \leq M(t)} Z_n(t) \geq c e^{\gamma(t-T(t))}\right) \leq \eta.$$

**Proof.** The proof is similar to the proof of Lemma 6.2. Let  $c > 0$  and define the processes  $(Z_n^*(t) : t \in \mathbb{R})$ , the sequence  $(A_n)$ , and the numbers  $\tau_n^*$  as above. We have

$$\{Z_n(t) \geq c e^{\gamma(t-T(t))}\} \subset \{Z_n^*(t) \geq c e^{\gamma(t-T(t))}\} \subset \{A_n \geq c e^{\gamma[(1-F_n)(t-T(t))-F_n(T(t)-\tau_n^*)]}\}.$$

Therefore, for any  $v > 1$  and  $n \geq vn(t)$ ,

$$\mathbb{P}\left(\max_{n \geq vn(t)} Z_n(t) \geq c e^{\gamma(t-T(t))}\right) \leq \mathbb{P}(T(t) \geq t/2) + \sum_{n=vn(t)}^{\infty} \mathbb{P}(A_n \geq c e^{\gamma[(1-F_n)^{t/2}-F_n(T(t)-\tau_n^*)]}). \quad (11)$$

An argument analogous to Lemma 6.2 yields, for any  $\delta > 0$ ,

$$\mathbb{P}\left(A_n \geq c e^{\gamma[(1-F_n)^{t/2} - F_n(T(t) - \tau_n^*)]}\right) \leq C_0 \mathbb{E}\left[\exp\left\{-\frac{c}{2} e^{\gamma[(1-F)^{t/2} - \delta(T(t) - \tau_n^*)]}\right\}\right] + C_0 \mathbb{P}(F < \delta),$$

where  $F$  is a random variable of law  $\mu$ . We now pick  $\varrho > \frac{\lambda^* + \varepsilon}{\gamma\delta}$ . As in Lemma 6.2 we use existence of a constant  $C_\varrho$  such that  $e^{-y} \leq C_\varrho y^{-\varrho}$ , for all  $y \geq 0$ , and Lemma 6.1 to get

$$\mathbb{E}\left[\exp\left\{-\frac{c}{2} e^{\gamma[(1-F)^{t/2} - \delta(T(t) - \tau_n^*)]}\right\}\right] \leq C_\varrho \left(\frac{c}{2}\right)^{-\varrho} e^{\gamma\varrho\varepsilon} \left(\frac{n(t)}{n}\right)^{\varrho \frac{\gamma\delta}{\lambda^* + \varepsilon}} \frac{K(0, \varrho\gamma/2)}{n(t)}.$$

Summing over  $n \geq vn(t)$  yields

$$\begin{aligned} \sum_{n \geq vn(t)} \mathbb{P}(A_n \geq c e^{\gamma[(1-F_n)^{t/2} - F_n(T(t) - \tau_n^*)]}) &\leq C_0 C_\varrho \left(\frac{c}{2}\right)^{-\varrho} e^{\gamma\varrho\varepsilon} \frac{K(0, \varrho\gamma/2)}{n(t)} \sum_{n=vn(t)}^{\infty} \left(\frac{n(t)}{n}\right)^{\varrho \frac{\gamma\delta}{\lambda^* + \varepsilon}} + C_0 \mathbb{P}(F < \delta) \\ &\leq C v^{1-\varrho \frac{\gamma\delta}{\lambda^* + \varepsilon}} + C_0 \mathbb{P}(F < \delta), \end{aligned}$$

where  $C$  is a constant that does not depend on  $t$  or  $v$ . Finally, using that  $1 - \varrho \frac{\gamma\delta}{\lambda^* + \varepsilon} < 0$ , recalling that  $\mathbb{P}(F < \delta) \rightarrow 0$ , as  $\delta \downarrow 0$ ,  $\mathbb{P}(T(t) \geq t/2) \rightarrow 0$ , as  $t \uparrow \infty$ , and plugging this into (11) completes the proof.  $\square$

### 6.3 Families born early are not fit enough

The following lemma is a standard extreme value result that is included for completeness.

**Lemma 6.4.** *For all  $\kappa, \eta > 0$ , there exists  $w = w(\kappa, \eta) > 0$  such that, for all  $t$  large enough,*

$$\mathbb{P}\left(\Gamma_t([- \infty, -\log w] \times [0, \kappa] \times [0, \infty]) = 0\right) \geq 1 - \eta.$$

**Proof.** We have

$$\begin{aligned} \mathbb{P}\left(\Gamma_t([- \infty, -\log w] \times [0, \kappa] \times [0, \infty]) = 0\right) &= \mathbb{P}(F_n < 1 - \kappa/t, \forall n \text{ s.t. } \tau_n \leq T(t) - \log w) \\ &= \mathbb{P}(F_n < 1 - \kappa/t, \forall n \leq M(T(t) - \log w)), \end{aligned}$$

where we recall that  $M(T(t) - \log w)$  is the number of families that were founded before time  $T(t) - \log w$ . Thus, in view of Hypothesis (RV),

$$\begin{aligned} \mathbb{P}\left(\Gamma_t([- \infty, -\log w] \times [0, \kappa] \times [0, \infty]) = 0\right) &= \mathbb{E}[\mu(0, 1 - \kappa/t)^{M(T(t) - \log w)}] \\ &= (1 + o(1)) \mathbb{E}\left[\exp(-M(T(t) - \log w)(\kappa/t)^\alpha \ell(\kappa/t))\right], \end{aligned}$$

when  $t$  tends to infinity. Note that, by Lemma 4.1, with probability tending to one,

$$\log w = T(t) - (T(t) - \log w) \leq \tau_{n(t)} - \tau_{M(T(t) - \log w)} \leq \frac{1}{\lambda^* - \varepsilon} \log \frac{n(t)}{M(T(t) - 1)} + \varepsilon,$$

implying that

$$M(T(t) - 1) \leq n(t) \exp[-(\lambda^* - \varepsilon)(\log w - \varepsilon)].$$

Recall that, using (RV) again,  $n(t) \sim t^\alpha / \ell(1/t)$ . We thus get

$$\mathbb{P}\left(\Gamma_t([- \infty, -\log w] \times [0, \kappa] \times [0, \infty]) = 0\right) \geq (1 + o(1)) \exp(-e^{-(\lambda^* - \varepsilon)(\log w - \varepsilon)} \kappa^\alpha \ell(\kappa/t) / \ell(1/t)).$$

Since  $\ell$  is a slowly varying function, we have that  $\ell(\kappa/t) / \ell(1/t) \rightarrow 1$  when  $t$  tends to infinity. In conclusion,

$$\mathbb{P}\left(\Gamma_t([- \infty, -\log w] \times [0, \kappa] \times [0, \infty]) = 0\right) \geq (1 + o(1)) \exp(-e^{-(\lambda^* - \varepsilon)(\log w - \varepsilon)} \kappa^\alpha) \rightarrow 1,$$

as  $w \uparrow \infty$ , which concludes the proof.  $\square$

## 7 Proof of non-extensiveness of condensation

### 7.1 Proof of Theorem 2.2

Let  $\eta, c > 0$ . By Lemma 6.2 there exists  $\kappa = \kappa(c, \eta)$  such that

$$\liminf_{t \rightarrow \infty} \mathbb{P}(\Gamma_t([-\infty, \infty] \times [\kappa, \infty] \times (c, \infty]) = 0) \geq 1 - \eta.$$

By Lemma 6.3 there exists  $v = v(c, \eta) > 1$  such that

$$\liminf_{t \rightarrow \infty} \mathbb{P}(\Gamma_t([0, \log v, \infty] \times [0, \infty] \times (c, \infty]) = 0) \geq 1 - \eta.$$

By Lemma 6.4, there exists  $w = w(\kappa, \eta) > 0$  such that

$$\liminf_{t \rightarrow \infty} \mathbb{P}(\Gamma_t([-\infty, -\log w] \times [0, \kappa] \times [0, \infty]) = 0) \geq 1 - \eta.$$

Finally, Proposition 5.1 gives that  $\Gamma_t$  converges on  $(-\infty, \log v) \times [0, \kappa] \times (c, \infty]$  to the Poisson process with intensity measure  $\zeta$ . Combining these four facts and using that  $\eta > 0$  is arbitrarily small, we get convergence on  $[-\infty, \infty] \times [0, \infty] \times (c, \infty]$ . As this holds for all  $c > 0$  the proof is complete.

### 7.2 Proof of Corollary 2.3

(i) We fix  $x > 0$  and apply the vague convergence proved in Theorem 2.2 to the compact set  $K := [-\infty, +\infty] \times [0, \infty] \times [x, \infty]$ . We get, as  $t \rightarrow \infty$ , that

$$\sum_{n=1}^{M(t)} \mathbf{1}_{K(\tau_n - T(t), (t - \tau_n)(1 - F_n), e^{-\gamma(t - T(t))} Z_n(t))} \Rightarrow \text{Poisson} \left( \int_K d\zeta \right).$$

Hence

$$\mathbb{P} \left( e^{-\gamma(t - T(t))} \max_{n \in \{1, \dots, M(t)\}} Z_n(t) \geq x \right) \rightarrow \mathbb{P} \left( \text{Poisson} \left( \int_K d\zeta \right) \geq 1 \right) = 1 - \exp \left( - \int_K d\zeta \right). \quad (12)$$

Integrating out gives

$$\begin{aligned} \int_K d\zeta &= \int_{-\infty}^{+\infty} \int_0^\infty \int_x^\infty \alpha f^{\alpha-1} \lambda^* e^{\lambda^* s} e^{-ze^{\gamma(s+f)}} e^{\gamma(s+f)} dz df ds \\ &= \int_0^\infty e^{-w} \int_0^\infty \alpha f^{\alpha-1} \int_{-\infty}^{\frac{1}{\gamma} \log \frac{w}{x} - f} \lambda^* e^{\lambda^* s} ds df dw \\ &= \left( \int_0^\infty e^{-w} \left( \frac{w}{x} \right)^{\frac{\lambda^*}{\gamma}} dw \right) \left( \int_0^\infty \alpha f^{\alpha-1} e^{-\lambda^* f} df \right) = \frac{\Gamma(\alpha+1) \Gamma(1 + \frac{\lambda^*}{\gamma})}{(\lambda^*)^\alpha} x^{-\frac{\lambda^*}{\gamma}}. \end{aligned}$$

Thus, the right hand side in (12) is  $1 - \exp(-\Lambda x^{-\eta})$ , for  $\Lambda = \frac{\Gamma(\alpha+1) \Gamma(1 + \frac{\lambda^*}{\gamma})}{(\lambda^*)^\alpha}$  and  $\eta = \frac{\lambda^*}{\gamma}$ . Summarising,

$$\mathbb{P} \left( \left( e^{-\gamma(t - T(t))} \max_{n \in \{1, \dots, M(t)\}} Z_n(t) \right)^{-\eta} \leq y \right) \rightarrow 1 - \exp(-\Lambda y) = \mathbb{P}(\text{Exp}(\Lambda) \leq y),$$

which proves the statement.

(ii) The probability that  $t(1 - V(t))$  is in an interval  $[a, b]$ , for some  $0 \leq a < b$ , converges to

$$\int_{-\infty}^{+\infty} \int_a^b \int_0^\infty e^{-\zeta([-\infty, +\infty] \times [0, \infty] \times [z, \infty])} \zeta(ds, df, dz),$$

where the inner integration is with respect to  $z$ , the middle with respect to  $f$ , and the outer with respect to  $s$ . We recall from above that

$$\zeta([-\infty, +\infty] \times [0, \infty] \times [z, \infty]) = \frac{\Gamma(\alpha+1)\Gamma(1+\frac{\lambda^*}{\gamma})}{(\lambda^*)^\alpha} z^{-\frac{\lambda^*}{\gamma}}.$$

Under (Cond), we have  $\lambda^* = \gamma$  and the right hand side becomes  $\frac{\alpha\Gamma(\alpha, \lambda^*)}{z}$ , where

$$\Gamma(\alpha, \lambda^*) := \int_0^\infty f^{\alpha-1} e^{-\lambda^* f} df = \frac{\Gamma(\alpha)}{(\lambda^*)^\alpha}.$$

We get, substituting  $v = e^{\gamma(s+f)}$  and recalling that  $\lambda^* = \gamma$ ,

$$\begin{aligned} \int_{-\infty}^{+\infty} \int_0^\infty e^{-\zeta([-\infty, +\infty] \times [0, \infty] \times [z, \infty])} d\zeta(s, f, z) &= \alpha f^{\alpha-1} e^{-\lambda^* f} df \int_0^\infty \left( \int_0^\infty v e^{-zv} dv \right) e^{-\alpha\Gamma(\alpha, \lambda^*)/z} dz \\ &= \alpha f^{\alpha-1} e^{-\lambda^* f} df \int_0^\infty \frac{e^{-\alpha\Gamma(\alpha, \lambda^*)/z}}{z^2} dz = \frac{f^{\alpha-1} e^{-\lambda^* f} df}{\Gamma(\alpha, \lambda^*)}. \end{aligned}$$

(iii) By Theorem 2.2 the random variable  $S(t) - T(t)$  converges to a random variable  $U$  with density

$$\int_0^\infty \int_0^\infty e^{-\zeta([-\infty, +\infty] \times [0, +\infty] \times [z, +\infty])} \zeta(s, df, dz).$$

### 7.3 Proof of Theorem 2.4

We have in view of Theorem 2.2, for all  $\varepsilon > 0$  as  $t \uparrow \infty$ ,

$$e^{-\gamma(t-T(t))} \sum_{n=1}^{M(t)} Z_n(t) = \int z d\Gamma_t(s, f, z) \geq \int z \mathbf{1}_{(\varepsilon, 1)}(z) d\Gamma_t(s, f, z) \rightarrow \int z \mathbf{1}_{(\varepsilon, 1)}(z) d\text{PPP}_\zeta(s, f, z),$$

where  $\text{PPP}_\zeta$  is the counting measure of a Poisson process with intensity measure  $\zeta$ . Observe that, for all  $m \in \mathbb{N}$ ,

$$\int z \mathbf{1}_{(\varepsilon, 1)}(z) d\text{PPP}_\zeta(s, f, z) \geq \sum_{k=0}^{m(1-\varepsilon)-1} \int_{\varepsilon + \frac{k}{m}}^{\varepsilon + \frac{k+1}{m}} z d\text{PPP}_\zeta(s, f, z) \geq \sum_{k=0}^{m(1-\varepsilon)-1} \left( \varepsilon + \frac{k}{m} \right) P_k,$$

where  $(P_k)_{k \geq 0}$  is a sequence of independent Poisson random variables of parameters  $\zeta(\mathbb{R} \times (0, \infty) \times (\varepsilon + \frac{k}{m}, \varepsilon + \frac{k+1}{m}))$ . As before, we find

$$\zeta(\mathbb{R} \times (0, \infty) \times (a, b)) = \frac{\Gamma(\alpha+1)\Gamma(1+\frac{\lambda^*}{\gamma})}{(\lambda^*)^\alpha} \left( a^{-\frac{\lambda^*}{\gamma}} - b^{-\frac{\lambda^*}{\gamma}} \right).$$

We abbreviate  $\Lambda := \frac{\Gamma(\alpha+1)\Gamma(1+\frac{\lambda^*}{\gamma})}{(\lambda^*)^\alpha}$  and get

$$\begin{aligned} \mathbb{E} \left[ \sum_{k=0}^{m(1-\varepsilon)-1} \left( \varepsilon + \frac{k}{m} \right) P_k \right] &= \Lambda \sum_{k=0}^{m(1-\varepsilon)-1} \left( \varepsilon + \frac{k}{m} \right) \left( \left( \varepsilon + \frac{k}{m} \right)^{-\frac{\lambda^*}{\gamma}} - \left( \varepsilon + \frac{k+1}{m} \right)^{-\frac{\lambda^*}{\gamma}} \right) \\ &\sim \frac{\Lambda}{m} \sum_{k=0}^{m(1-\varepsilon)-1} \left( \varepsilon + \frac{k+1}{m} \right)^{-\frac{\lambda^*}{\gamma}} \sim \Lambda \int_0^{1-\varepsilon} (\varepsilon + x)^{-\frac{\lambda^*}{\gamma}} dx, \end{aligned}$$

as  $m \rightarrow \infty$ , by Riemann integration. The right hand side is of order  $\log(1/\varepsilon)$  if  $\lambda^* = \gamma$ , and of order  $\varepsilon^{1-\frac{\lambda^*}{\gamma}}$  if  $\lambda^* > \gamma$ . In any case, the expectation goes to infinity, as  $\varepsilon \downarrow 0$ . With a similar reasoning we get

$$\text{Var} \left[ \sum_{k=0}^{m(1-\varepsilon)-1} \left( \varepsilon + \frac{k}{m} \right) P_k \right] \leq \frac{\Lambda}{m} \sum_{k=0}^{m(1-\varepsilon)-1} \frac{(\varepsilon + \frac{k}{m})^{2-\frac{\lambda^*}{\gamma}}}{\varepsilon + \frac{k+1}{m}} \sim \Lambda \int_0^{1-\varepsilon} (\varepsilon + x)^{1-\frac{\lambda^*}{\gamma}} dx.$$

If  $\lambda^* = \gamma$  the variance is therefore bounded, and otherwise it grows of a slower order than the square of the expectation. Thus, by Chebyshev's inequality, we get

$$\lim_{\varepsilon \downarrow 0} \lim_{m \rightarrow \infty} \sum_{k=0}^{m(1-\varepsilon)-1} \left( \varepsilon + \frac{k}{m} \right) P_k = \infty,$$

in probability, and this implies the claimed result.

## 8 Open problems

### Precise growth of the system

A question that remains open is about the precise growth of  $N(t)$  in the condensation phase. Recall from Remark 1 that if condensation is absent we have  $\log N(t) = \lambda^* t + O(1)$  but we do not have a similarly strong statement in the condensation case. We get a lower bound on the growth by considering the size of the largest family. Under (RV) this gives  $\log N(t) - \gamma t + \gamma T(t) \rightarrow \infty$  with  $T(t) \sim \alpha/\lambda^* \log t$ . A plausible conjecture would be that in the condensation case this bound is sharp to logarithmic order, i.e.  $\log N(t) = \lambda^* t - \alpha \log t + o(\log t)$ .

### Shape of the condensate

Our results offer only a partial answer to the question raised in Borgs et al. [7] how the links in the network are distributed among the highest fitnesses present in the system at any given time. The most important question that remains open here is whether the families that together form the condensate have a characteristic collective behaviour prior to condensation. The work on Kingman's model in Dereich and Mörters [12], and on a growth model without self-organisation in Dereich [10], suggests that this is indeed the case. In particular in the model of [10] it is shown that for parameters chosen in the condensation regime, the random mass distribution in a suitably shrinking neighbourhood of the maximal fitness value satisfies a law of large numbers with limiting shape given by a Gamma distribution. We believe that this is a phenomenon of universal nature and conjecture the same behaviour in our model.

**Conjecture 8.1** (Condensation wave). *Under assumption (RV) we have*

$$\lim_{t \rightarrow \infty} \Xi_t(1 - \frac{x}{t}, 1) = \frac{\omega(\beta, \gamma)}{\Gamma(\alpha + 1)} \int_0^x y^\alpha e^{-y} dy,$$

*in probability, i.e. the condensation wave has the shape of a Gamma distribution with shape parameter  $1 + \alpha$ .*

### Other classes of bounded fitness distributions

In this paper we have investigated the class of fitness distributions in the maximal domain of attraction of the Weibull distribution, i.e. those bounded distributions of regular variation at the maximal fitness value. It would also be interesting to discuss fitness distributions with a faster decay at the maximal fitness value, for example in the maximal domain of attraction of the Gumbel distribution. This includes the interesting example of distributions with  $\log \mu(1 - \varepsilon, 1) \sim -\varepsilon^{-\gamma}$ , for some  $\gamma > 0$ . What is the shape of the condensation wave in this case? Will we also see non-extensive condensation? More generally, can we find bounded fitness distributions where we experience condensation by macroscopic occupancy? This circle of problems is currently under investigation.

**Acknowledgements:** We are grateful to EPSRC for support through grant EP/K016075/1, and SFB 878 as well as the *Institut für mathematische Statistik* at the University of Münster for hosting PM during a sabbatical in 2015/16. We also thank Achim Klenke for useful discussions on branching processes with selection and mutation, which were the starting point of this project. Thanks are also due to Anna Senkevich for providing insightful simulations, and in particular the figure in this paper. Last but not least we would like to thank the organisers of the workshop *Interplay of Probability and Analysis in Applied Mathematics* at the Mathematisches Forschungsinstitut Oberwolfach in 2015, where this work was presented and extensively discussed.

## References

- [1] Krishna B. ATHREYA and Samuel KARLIN. Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. *Ann. Math. Statist.*, 39 (1968) 1801-1817.
- [2] Krishna B. ATHREYA and Peter E. NEY. Branching Processes. *Springer-Verlag*, 1972.
- [3] Michiel VAN DEN BERG, John T. LEWIS, and Joe PULÉ. A general theory of Bose-Einstein condensation. *Helvetica Physica Acta* 59 (1986) 1271-1288.
- [4] Ginestra BIANCONI and Albert-Laszlo BARABÁSI. Bose-Einstein condensation in complex networks. *Physical Review Letters*, 86 (2001) 5632-5635.
- [5] Nick BINGHAM, Charles GOLDIE, and Jozef TEUGELS. Regular Variation. *Cambridge University Press*, 1989.
- [6] Shankar BHAMIDI. Universal techniques to analyze preferential attachment trees: Global and local analysis. Preprint available at <http://www.unc.edu/~bhamidi/>.
- [7] Christian BORGS, Jennifer CHAYES, Constantinos DASKALAKIS and Sebastien ROCH. First to market is not everything: An analysis of preferential attachment with fitness. In STOC'07: Proceedings of the 39th Annual ACM Symposium on Theory of Computing, ACM, pp. 135-144, see [arXiv:0710.4982](https://arxiv.org/abs/0710.4982) for a long version.
- [8] Fan CHUNG, Shirin HANDJANI and Doug JUNGREIS. Generalizations of Pólya's urn problem. *Annals of Combinatorics*, 7 (2003) 141-153.
- [9] Andrea COLLEVECCHIO, Codina COTAR, and Marco LICALZI. On a preferential attachment and generalized Pólya's urn model. *Annals of Applied Probability*, 23 (2013) 1219-1253.
- [10] Steffen DEREICH. Preferential attachment with fitness: Unfolding the condensate. *Electronic Journal of Probability*, 21 (2016), Paper No. 3, 38 pp.
- [11] Steffen DEREICH, Cécile MAILLER, and Peter MÖRTERS. Traveling waves in reinforced branching processes. *In preparation*.
- [12] Steffen DEREICH and Peter MÖRTERS. Emergence of condensation in Kingman's model of selection and mutation. *Acta Appl. Math.*, 127 (2013) 17-26.
- [13] Steffen DEREICH and Marcel ORTGIESE. Robust analysis of preferential attachment models with fitness. *Combin. Probab. Comput.* 23 (2014) 386-411.
- [14] Claude GODRÈCHE and Jean Marc LUCK. On leaders and condensates in a growing network. *Journal of Statistical Mechanics: Theory and Experiment* 07 (2010) P07031.
- [15] Andrea HODGINS-DAVIS, Daniel P. RICE and Jeffrey P. TOWNSEND. Gene expression evolves under a house-of-cards model of stabilizing selection. *Mol. Biol. Evol.* (2015) doi: 10.1093/molbev/msv094.
- [16] Svante JANSON. Functional limit theorems for multitype branching processes and generalized Polya urns. *Stochastic Processes and their Applications* 110 (2004) 177-245.
- [17] Svante JANSON. Simply generated trees, conditioned Galton-Watson trees, random allocations and condensation. *Probability Surveys*, 9 (2012) 103-252.
- [18] John KINGMAN, A simple model for the balance between selection and mutation. *Journal Appl. Prob.* 15 (1978) 1-12.
- [19] Júlia KOMJÁTHY, Explosive Crump-Mode-Jagers branching processes. *Preprint arXiv:1602.01657*.
- [20] Olle NERMAN. On the convergence of supercritical general (C-M-J) branching processes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 57 (1981) 365-395.
- [21] Robin PEMANTLE. A survey of random processes with reinforcement. *Probability Surveys*, 4 (2007) 1-79.
- [22] Sidney RESNICK. Extreme values, regular variation, and point processes. *Springer-Verlag*, 1987.